

Trust Dynamics in Multi-Agent Coalition Formation

Dariusz G. Mikulski ^{*a,c}, Frank L. Lewis ^b, Edward Y. Gu ^c, Greg R. Hudas ^a

^a Ground Vehicle Robotics (GVR), U.S. Army RDECOM-TARDEC, Warren, MI, USA

^b Automation & Robotics Research Institute, University of Texas at Arlington, Fort Worth, TX, USA

^c Department of Electrical and Computer Engineering, Oakland University, Rochester, MI, USA

ABSTRACT

We present a rigorous treatment of coalition formation based on trust interactions in multi-agent systems. Current literature on trust in multi-agent systems primarily deals with trust models and protocols of interaction in non-cooperative scenarios. Here, we use cooperative game theory as the underlying mathematical framework to study the trust dynamics between agents as a result of their trust synergy and trust liability in cooperative coalitions. We rigorously justify the behaviors of agents for different classes of games, and discuss ways to exploit the formal properties of these games for specific applications, such as unmanned cooperative control.

Keywords: Trust, Multi-Agent Systems, Cooperative Game Theory, Coalitions, Convoys, Unmanned Military Systems

1. INTRODUCTION

The concept of trust is something people intuitively understand, but might find difficultly grasping rationally. People rely on trust whenever they need to gauge something they cannot ever know precisely with reasonable time or effort. But this reliance exposes them to vulnerabilities associated with betrayal. So, on one hand, trust can be used as highly versatile heuristic that deals with uncertainty by reducing the complexity of expectations within arbitrary situations involving risk, vulnerability, and interdependence¹. On the other hand, the motivation for trust – the need to believe that things will behave consistently – exposes individuals to potentially undesirable outcomes.

Trust is clearly a “double-edged sword,” but its importance in military contexts is evident. Currently, the United States military wages asymmetric battles against insurgencies in Southwest Asia, where enemy combatants have exhibited quick and deadly adaptations to US strategies and tactics². One of these deadly adaptations has been the use of improvised explosive devices (IEDs). Early in the Iraq and Afghanistan war efforts, IEDs were jury-rigged homemade bombs that, while deadly, could be avoided with increased awareness. But insurgents quickly adapted by developing more sophisticated explosives, often with timing devices, pressure switches, and even wireless triggers. In addition, insurgents became more difficult to detect due to their knowledge of the local terrain and their ability to mix with civilian populations. Responses to more advanced IED attacks required Soldiers to put more of their trust into new equipment (such as up-armor vehicles, electronic jammers, and robots) as well as local allies. And while this did not imply that any Soldier was any safer than before, the trust helped Soldiers deal internally with wartime uncertainties so that they could continue their duties and focus on mission objectives.

The need for trust in extreme military situations is obvious; but what may not be as obvious is the effect of trust on more ordinary military interactions. Trust impacts a range of social processes between Soldiers that influence the cognitive and physical strain of being at war. When the trust of a Soldier becomes lower toward other people, equipment, or processes, then he will likely need to exert more effort in order to resolve any perceived uncertainties. This extra effort could manifest itself into a distraction that lowers the effectiveness of the Soldier at best. However, in prolonged high stress situations, this additional effort could also manifest itself as a persistently guarded psychological state that monitors for violations of expectations and predictions.

* Dariusz G. Mikulski . Email: dariusz.mikulski@us.army.mil .

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 25 APR 2011		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Trust Dynamics in Multi-Agent Coalition Formation			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Dariusa Mikulski; Frank L. Lewis; Edward Y. Gu; Greg R. Hudas			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army RDECOM-TARDEC 6501 E 11 Mile Rd Warren, MI 48397-5000, USA Automation & Robotics Research Institute, University of Texas at Arlington, Forth Worth, TX, USA Department of Electrical and Computer Engineering, Oakland University, Rochester, MI USA			8. PERFORMING ORGANIZATION REPORT NUMBER 21562		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Army RDECOM-TARDEC 6501 E 11 Mile Rd Warren, MI 48397-5000, USA			10. SPONSOR/MONITOR'S ACRONYM(S) TACOM/TARDEC/RDECOM		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 21562		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES Presented at SPIE 25-29 April 2011 Orlando, Florida, USA, The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 14	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Trust can, therefore, be seen as critical survival tool for Soldiers dealing with uncertainty, whether in combat or not. The effect of high trust in military settings can lessen the defensive monitoring of others, reduce in the need for hierarchical control, improve cooperation due to increased predictability and expectations of reciprocity, improve information sharing (with less need to filter unfavorable information), lower levels of conflict (friction and dissent), and improve group performance and processes³. Our goal here then is to determine a way to incorporate elements of trust into machine-based agents in an effort to realize similar emergent efficiencies, such as lower power consumption, faster algorithms, and the exchange of higher-quality information. We hypothesize that machines with an ability to accurately evaluate trustworthiness in other agents will be able to use this knowledge to intelligently select better partners for cooperative activities and achieve greater cooperative payoffs over time. Therefore, in this paper, we develop a theoretical framework for coalition formation based on trust interactions in multi-agent systems. Then, we show how the theory could be applied to a military unmanned system – specifically the unmanned military convoy.

2. REVIEW OF MULTI-AGENT TRUST RESEARCH

This section briefly highlights prior work in multi-agent trust research and intends to present a broad range of possible multi-agent trust approaches. For an extensive survey of trust research in multi-agent systems up to 2004, we urge the reader to examine the paper by Ramchurn, Huynh, and Jennings⁴. We borrow their classification of multi-agent trust approaches – trust models and protocols of interaction – to present both long-standing and recent findings.

2.1 Trust Models

Trust models give agents the ability to reason about the reciprocity, honesty, and reliability of other agents. Since agents in a system are always assumed to have selfish interests, these models take the viewpoint of an agent trying to find the most reliable interaction agents from a pool of potential agents.

Some trust model research attempts to characterize trust within non-cooperative scenarios. Sen demonstrates how reciprocity can emerge when agents learn to predict the value of future benefits when competitive agents cooperate⁵. Mukherjee, Banerjee, and Sen show how trust can be acquired if agents know their opponents chosen move in advance⁶. Castelfranchi and Falcone assert that socio-cognitive models which incorporate beliefs in competence, willingness, persistence, and motivation are essential to determine the amount of trust each agent can place in other agents^{7,8,9}.

Other work in trust models factor in evidence to justify trust values. Witkowski, Artikis, and Pitt propose a model whereby trust is based on performance in past interactions¹⁰. Sabater and Sierra, through the REGRET system, attribute fuzziness to the notion of performance and adopt a sociological approach to reputation by using a weighted sum of subjective impressions¹¹. Teacy, Patel, Jennings, and Luck develop a probabilistic trust model in terms of confidence that expected values lie within specific error tolerances¹². Theodorakopoulos and Baras focus on evaluating trust evidence in ad hoc networks using the theory of semirings¹³. Wang and Singh define trust in terms of belief and certainty, and formulate certainty in terms of evidence based on a statistical measure defined over a probability distribution of positive outcome probabilities¹⁴.

Some work also incorporates trust models into specific applications. Abdul-Rahman and Hailes attempt to use social trust characteristics and word-of-mouth to calculate trust in virtual environments¹⁵. Zhang, Das, and Liu present a framework to secure data aggregation against false data injection in wireless sensor networks that exploits redundancy in gathered data to evaluate the trustworthiness of each sensor¹⁶. Ballal and Lewis discuss the concept of trust consensus for collaborative control and show how the propagation of trust through a network can lead to a global asymptotic trust consensus among all agents¹⁷.

2.2 Protocols of Interaction

Whereas trust models are intended to build trust at the agent-level, protocols of interaction are intended to build trust at the system-level. In short, they are developed to make sure agents will gain some utility if they follow the rules – and lose utility if they don't. Thus, the rules of a system enable an agent to trust other agents by the virtue of the different constraints in a system.

Multi-agent trust protocols can be divided into three main groups: truth-eliciting, reputation mechanisms, and security mechanisms. Truth-eliciting protocols force agents to follow the rules, which dictate the individual steps in interactions and the information revealed by the agents during interactions. By doing so, agents should find no better option than telling the truth. The Vickrey-Clarke-Groves (VCG) mechanisms are an example of protocols that enforce truth-telling¹⁸. Reputation mechanisms force agents to interact with some trusted authority to get public ratings on other

agents in a system. Zacharia and Maes outlined some basic requirements for practical reputation mechanisms¹⁹. For security in agent networks, trust is used to describe the fact that an agent can prove who they say they are. Poslad, Calisti, and Charlton proposed that identity, access permissions, content integrity, and content privacy are essential for agents to trust each other and each other's messages transmitted across a network²⁰. These requirements are specified in the Foundation for Intelligent Physical Agents (FIPA) abstract architecture and implemented by public key encryption (PGP and X.509) and a certificate infrastructure²¹.

3. CLASSES OF TRUST GAMES

This section characterizes different classes of trust games within the context of cooperative game theory. Our characterizations provide the necessary conditions for a coalition trust game to be classified into a particular class. We start with additive and constant-sum trust games, which have limited value for cooperative applications, but are included for completeness. Then, we discuss superadditive and convex trust games, and show the necessary conditions for agents to form a grand coalition. In general, grand coalition solution concepts presented here can also be applied to smaller coalitions within a trust game through the use of a trust subgame.

3.1 Preliminaries

Cooperative game theory focuses on what groups of self-interested agents can achieve. It is not concerned with how agents make choices or coordinate in coalitions, and does not assume that agents will always agree to follow arbitrary instructions. Rather, cooperative game theory defines games that tell how well a coalition can do for itself²². And while the coalition is the basic modeling unit for coalition game, the theory supports modeling individual agent preferences without concern for their possible actions. As such, it is an ideal framework for modeling trust-based coalition formation since it can show how each agent's trust preferences can influence a group's ability to reason about trustworthiness. In essence, cooperative game theory allows us to blend the reasoning abilities of agent-level trust models with the system-level benefits of trust interaction protocols, forming a general concept of *system-level reasoning*.

Definition: Let $\Gamma = (N, v)$ be a coalitional trust game with transferable utility where:

- N is a finite set of agents, indexed by i
- $v: 2^N \rightarrow \mathbb{R}$ associates with each coalition $S \subseteq N$ a real-valued payoff $v(S)$ that is distributed between the agents. Singleton coalitions, by definition, are assigned no value; i.e. $v(i) = 0 \quad \forall i \in N$.

The transferable utility assumption means that payoffs in a coalition may be freely distributed among its members. With regards to payoff value of trust between agents, this assumption can be interpreted as a universal means for agents to mutually share the value of their trustworthy relationships. Trust cultivation often requires reciprocity between two agents as a necessary behavior to develop trust, and a transferable utility is a convenient way to model the exchange for this notion.

In defining a transferable payoff value of trust, one aspect to consider are the "goods of trust". These refer to opportunities for cooperative activity, knowledge, and autonomy. In this paper, we refer to these goods as *trust synergy* $s(S)$, which is a trust-based result that could not be obtained independently by two or more agents. We may also interpret trust synergy as the value obtained by agents in a coalition as a result of being able to work together due to their attitudes of trust for each other. In defining a set function for trust synergy, it is important to explicitly show how each agent's attitude of trustworthiness for every other agent in a coalition affects this synergy. In general, higher levels of trust in a coalition should produce higher levels of synergy.

The payoff value of trust, however, also includes an opposing force in the form of vulnerability exposure, which we refer to as *trust liability* $l(S)$. Trusting involves being optimistic that the trustee will do something for the truster; and this optimism is what causes the vulnerability, since it restricts the inferences a truster makes about the likely actions of the trustee. However, the refusal to be vulnerable tends to undermine trust since it does not allow others to prove their own trustworthiness, stifling growth in trust synergy. Thus, we see that agents in trust-based relationships with other agents must be aware of the balance between the values of the trust synergy and trust liability in addition to their relative magnitudes.

Definition: Let the characteristic payoff function of a trust game be the difference between the trust synergy and trust liability of a coalition S .

$$v(S) = s(S) - l(S) \quad (1)$$

This payoff is similar to the well-known *constrained coalitional game (CCG)* that incorporates gains from cooperation with the costs due to communications network restrictions²³. However, the characteristic function v in CCGs is defined on the structure of a particular communications network between agents, whereas the characteristic function for our trust game is defined only on a set of agents. As such, agents who are completely disconnected from communication with other agents can still theoretically maintain membership in the same trust-based coalition.

3.2 Additive Trust Game

Additive games are considered inessential games in cooperative game theory since the value of the union of two disjoint coalitions ($S_1 \cap S_2 = \emptyset$) is equivalent to the sum of the values of each coalition.

$$v(S_1 \cup S_2) = v(S_1) + v(S_2) \quad \forall S_1, S_2 \subset N \quad (2)$$

In (2), we see that the total value of the trust relationships *between* any two disjoint coalitions must always be zero. In other words, the trust synergy between any two disjoint coalitions must always result in a value that is equal to their trust liability. Thus, by expanding this definition for trust games and rearranging the terms, we can characterize an additive trust game as:

$$s(S_1 \cup S_2) - l(S_1 \cup S_2) = s(S_1) - l(S_1) + s(S_2) - l(S_2) \quad \{\forall S_1, S_2 \subset N: S_1 \cap S_2 = \emptyset\} \quad (3)$$

$$s(S_1 \cup S_2) - s(S_1) - s(S_2) = l(S_1 \cup S_2) - l(S_1) - l(S_2) \quad \{\forall S_1, S_2 \subset N: S_1 \cap S_2 = \emptyset\} \quad (4)$$

3.3 Constant-Sum Trust Game

In constant-sum games, the sum of all coalition values in N remains the same, regardless of any outcome.

$$v(N) = v(S) + v(N \setminus S) = k \quad \forall S \subset N \quad (5)$$

By expanding this definition for trust games and rearranging the terms, we can see that the constant-sum trust game is a special case of a two-coalition additive trust game involving every agent in the game.

$$s(N) - l(N) = s(S) - l(S) + s(N \setminus S) - l(N \setminus S) \quad \forall S \subset N \quad (6)$$

$$s(N) - s(S) - s(N \setminus S) = l(N) - l(S) - l(N \setminus S) \quad \forall S \subset N \quad (7)$$

Definition: An agent is a *dummy agent* if the amount the agent contributes to any coalition is exactly the amount that it is able to achieve alone.

Theorem 1: Γ is a constant-sum trust game implies that Γ is a zero-sum trust game.

Proof: If Γ is a constant-sum game, then by (7), the following constraint for singleton coalitions must always hold:

$$s(N) - s(i) - s(N \setminus i) = l(N) - l(i) - l(N \setminus i) \quad \forall i \in N \quad (8)$$

By rearranging the terms, combining, and substituting, we get:

$$s(N) - l(N) = s(i) - l(i) + s(N \setminus i) - l(N \setminus i) \quad \forall i \in N \quad (9)$$

$$v(N) = v(i) + v(N \setminus i) \quad \forall i \in N \quad (10)$$

$$v(N) = v(N \setminus i) \quad \forall i \in N \quad (11)$$

The result in (11) implies that every agent in N must behave like a dummy agent if Γ is a constant-sum trust game. Since all agents behave like dummy agents and $v(i) = 0$ for all $i \in N$, then any coalition that forms in Γ will have no value. Hence, the value of the grand coalition is zero (i.e. $v(N) = k = 0$). Therefore, the only possible constant-sum trust game is the zero-sum trust game. This completes the proof.

Corollary 1: Γ is a zero-sum trust game if $s(S) = l(S) \ \forall S \subset N$.

Proof: If $s(S) = l(S) \ \forall S \subset N$, then $v(S) = 0 \ \forall S \subset N$. Thus, by (5), $v(N) = v(N \setminus S) = k \ \forall S \subset N$. This result implies that every possible coalition in N must behave like a coalition of dummy agents in a constant-sum trust game and their combinations with other coalitions will yield no value. Hence, the value of the grand coalition is always zero (i.e. $v(N) = k = 0$). This completes the proof.

Our proofs show that any constant-sum trust game is necessarily a zero-sum trust game that represents a special case of an additive trust game. These facts reinforce a notion that a group of agents who do not trust each other will always prefer to work as singleton coalitions. And even if there is some mutual trust between agents, gains from trust synergy are always lost to the trust liability, making it irrational to form any coalition with any other agent. Thus, if one determines that Γ is a constant-sum trust game, then this provides immediate justification for using non-cooperative game theory as the basis for modeling the purely competitive agents within the game since there is no payoff from cooperative activity.

3.4 Superadditive Trust Game

In a superadditive game, the value of the union of two disjoint coalitions ($S_1 \cap S_2 = \emptyset$) is never less than the sum of the values of each coalition.

$$v(S_1 \cup S_2) \geq v(S_1) + v(S_2) \ \forall S_1, S_2 \subset N \quad (12)$$

This implies a monotonic increase in the value of any coalition as the coalition gets larger.

$$S \subseteq A \subseteq N \rightarrow v(S) \leq v(A) \leq v(N) \quad (13)$$

This property of superadditivity tells us that the new links that are established between the agents in the two disjoint coalitions are the sources of the monotonic increases. This results in a snowball effect that causes all agents in the game to form the grand coalition (a coalition containing all agents in the game) since the total value of the new trust relationships *between* any two disjoint coalitions must always be positive semi-definite. In other words, the trust synergy between any two disjoint coalitions must always result in a value that is at least as large as their trust liability. Thus, by expanding (12) for trust games and rearranging the terms, we can characterize a superadditive trust game as:

$$s(S_1 \cup S_2) - l(S_1 \cup S_2) \geq s(S_1) - l(S_1) + s(S_2) - l(S_2) \ \{\forall S_1, S_2 \subset N: S_1 \cap S_2 = \emptyset\} \quad (14)$$

$$s(S_1 \cup S_2) - s(S_1) - s(S_2) \geq l(S_1 \cup S_2) - l(S_1) - l(S_2) \ \{\forall S_1, S_2 \subset N: S_1 \cap S_2 = \emptyset\} \quad (15)$$

3.5 Convex Trust Game

A game is convex if it is supermodular, and this trivially implies superadditivity (when $S_1 \cap S_2 = \emptyset$). Thus, we see that convexity is a stronger condition than superadditivity since the restriction that two coalitions must be disjoint no longer applies.

$$v(S_1 \cup S_2) + v(S_1 \cap S_2) \geq v(S_1) + v(S_2) \ \forall S_1, S_2 \subset N \quad (16)$$

In convex games, the incentives of joining a coalition grow as the coalition gets larger. This means that the marginal contribution of each agent $i \in N$ is non-decreasing.

$$v(S \cup i) - v(S) \leq v(A \cup i) - v(A) \text{ whenever } S \subset A \subset N \setminus i \quad (17)$$

Definition: A *subgame* $v_R: 2^R \rightarrow \mathbb{R}$, where $R \subseteq N$ is not empty, is defined as $v_R(S) = v(S)$ for each $S \subseteq R$. In general, solution concepts that apply to a grand coalition can also apply to smaller coalitions in terms of a subgame.

Definition: Given a game $\Gamma = (N, v)$ and a coalition $R \subseteq N$, the R -marginal game $v_R: 2^{N \setminus R} \rightarrow \mathbb{R}$ is defined by $v_R(S) = v(R \cup S) - v(R)$ for each $S \subseteq N \setminus R$.

Using these definitions, Branzei, Dimitrov, and Tijs proved that a game is convex if and only all of its marginal games are superadditive²⁴. We provide their proof here as a means for the reader to readily justify this assertion.

Theorem 2: A game $\Gamma = (N, v)$ is convex if and only if for each $R \in 2^N$ the R -marginal game $(N \setminus R, v_R)$ is superadditive.

Proof:

(i) Suppose (N, v) is convex. Let $R \subseteq N$ and $S_1, S_2 \subseteq N \setminus R$. Then:

$$\begin{aligned}
 & v_R(S_1 \cup S_2) + v_R(S_1 \cap S_2) \\
 &= v(R \cup S_1 \cup S_2) + v(R \cup (S_1 \cap S_2)) - 2v(R) \\
 &= v((R \cup S_1) \cup (R \cup S_2)) + v((R \cup S_1) \cap (R \cup S_2)) - 2v(R) \\
 &\geq v(R \cup S_1) + v(R \cup S_2) - 2v(R) \\
 &= (v(R \cup S_1) - v(R)) + (v(R \cup S_2) - v(R)) \\
 &= v_R(S_1) + v_R(S_2)
 \end{aligned} \tag{18}$$

where the inequality follows from the convexity of v . Hence, v_R is convex (and superadditive as well).

(ii) Let $S_1, S_2 \subseteq N$ and $R = S_1 \cap S_2$. Suppose that for each $R \in 2^N$, the game $(N \setminus R, v_R)$ is superadditive. If $R = \emptyset$, then the game $(N \setminus \emptyset, v_\emptyset) = (N, v)$ and $v(\emptyset) = 0$; hence, Γ is superadditive. If $R \neq \emptyset$, then because $(N \setminus R, v_R)$ is superadditive:

$$v_R((S_1 \cup S_2) \setminus R) \geq v_R(S_1 \setminus R) + v_R(S_2 \setminus R) \tag{20}$$

$$v(S_1 \cup S_2) - v(R) \geq v(S_1) - v(R) + v(S_2) - v(R) \tag{21}$$

$$v(S_1 \cup S_2) + v(R) \geq v(S_1) + v(S_2) \tag{22}$$

$$v(S_1 \cup S_2) + v(S_1 \cap S_2) \geq v(S_1) + v(S_2) \tag{23}$$

This completes the proof.

By using this characterization in Theorem 2 and expanding it to our definition of a trust game, we can state a necessary requirement to produce a convex trust game: that the marginal trust synergy between any two coalitions must always result in a value that is at least as large as their marginal trust liability.

$$s_R((S_1 \cup S_2) \setminus R) - l_R((S_1 \cup S_2) \setminus R) \geq s_R(S_1 \setminus R) - l_R(S_1 \setminus R) + s_R(S_2 \setminus R) - l_R(S_2 \setminus R) \tag{24}$$

$$\{\forall S_1, S_2 \subset N: S_1 \cap S_2 = R\}$$

$$s_R((S_1 \cup S_2) \setminus R) - s_R(S_1 \setminus R) - s_R(S_2 \setminus R) \geq l_R((S_1 \cup S_2) \setminus R) - l_R(S_1 \setminus R) - l_R(S_2 \setminus R) \tag{25}$$

$$\{\forall S_1, S_2 \subset N: S_1 \cap S_2 = R\}$$

Convex games are convenient due to several nice, well-known properties.

- The core of a convex game is never empty.
- Convex games are totally balanced, meaning that their subgames are also convex, each with a non-empty core.

- Convex games have a stable set that coincide with its core.
- The Shapley value of a convex game is the barycenter of the core.
- The vertices of a core can be found in polynomial time using a polyhedron greedy algorithm²⁵.

4. A PRACTICAL MODEL FOR TRUST GAMES

In the previous section, we characterized different classes of trust games without explicitly defining a trust game model. In this section, we provide a general model for trust games that conforms to the theoretical constructions in the previous section and can be adapted to a wide variety of applications.

4.1 Managing Agent Trust Preferences

The attitude of trustworthiness agents have toward other agents in a trust game is managed in an $|N| \times |N|$ matrix T .

$$T = [t_{i,j}]_{|N| \times |N|} = \begin{cases} t_{i,j} = 1, & i = j \\ t_{i,j} \in [0,1], & i \neq j \end{cases} \quad (26)$$

This matrix is populated with values $t_{i,j}$ that represent the probability that agent j is trustworthy from the perspective of agent i . The values $t_{i,j}$ can also be interpreted as the probabilities that agent i will allow agent j to interact with him, since rational agents would prefer to interact with more trustworthy agents.

The manner in which $t_{i,j}$ is evaluated depends on an underlying trust model. We make no assumption about the use of a particular trust model, as the choice of an appropriate model may be application-specific. We also make no assumption about the spatial distribution of the agents in a game – therefore, this matrix should not necessarily imply the structure of a communications graph.

4.2 Modeling Trust Synergy and Trust Liability

We provide a general model for trust synergy and trust liability that can be adapted for a variety of applications. Our model makes use of a symmetric matrix Σ to manage potential trust synergy and a matrix Λ to manage potential trust liability. Σ is symmetric because we assume that agents mutually agree on the benefits of a synergetic interaction.

$$\Sigma = [\sigma_{i,j}]_{|N| \times |N|} = \begin{cases} \sigma_{i,j} = 0, & i = j \\ \sigma_{i,j} = \sigma_{j,i} \geq 0, & i \neq j \end{cases} \quad (27)$$

$$\Lambda = [\lambda_{i,j}]_{|N| \times |N|} = \begin{cases} \lambda_{i,j} = 0, & i = j \\ \lambda_{i,j} \geq 0, & i \neq j \end{cases} \quad (28)$$

As with the T matrix, we make no assumptions about how Σ and Λ are calculated, since the meaning of their values may depend on the application. For example, the calculations for $\sigma_{i,j}$ and $\lambda_{i,j}$ between two agents may not only take into account each agent's individual intrinsic attributes – it may also factor in externalities (i.e. political climate, weather conditions, pre-existing conditions, etc.) that neither agent has direct control over.

Definition: The total value of the trust synergy in a coalition is defined as the following set function:

$$s(S) = \sum_{i,j \in S} \sigma_{i,j} t_{i,j} t_{j,i} \quad \forall i > j \quad (29)$$

Trust synergy is the value obtained by agents in a coalition as a result of being able to work together due to their attitudes of trust for each other. The set function $s(S)$ assumes that the events “agent i allows agent j to interact” and “agent j allows agent i to interact” are independent. This is reasonable since agents are assumed to behave as independent entities within a trust game (i.e. no agent is controlled by any other agent). Therefore, we treat the product $t_{i,j} t_{j,i}$ as the relative strength of a trust-based synergetic interaction, which justifies the use of the summation. The value for $\sigma_{i,j}$ serves as a weight for a trust-based synergetic interaction.

Definition: The total value of the trust liability in a coalition is defined as the following set function:

$$l(S) = \sum_{i,j \in S} \lambda_{i,j} t_{i,j} \quad \forall i \neq j \quad (30)$$

Trust liability can be thought of as the vulnerability that agents in a coalition expose themselves to due to their attitudes of trust for each other. We treat the product $\lambda_{i,j} t_{i,j}$ as a measure for agent i 's exposure to unfavorable trust-based interactions from agent j . A high amount of trust can expose agents to high levels of vulnerability. But each agent can regulate its exposure to trust liability by adjusting $t_{i,j}$. Changes to $t_{i,j}$, however, also influence the benefits of trust synergy.

4.3 Modeling the Trust Game

Definition: From (1), we define the trust game (also known as the total value of the trust payoff in a coalition) as the difference between its trust synergy and trust liability.

$$v(S) = \sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j} t_{i,j} t_{j,i} - \sum_{\substack{i,j \in S \\ \forall i \neq j}} \lambda_{i,j} t_{i,j} \quad (31)$$

$$v(S) = \sum_{\substack{i,j \in S \\ \forall i > j}} t_{i,j} t_{j,i} \left(\sigma_{i,j} - \frac{\lambda_{i,j}}{t_{j,i}} - \frac{\lambda_{j,i}}{t_{i,j}} \right) \quad (32)$$

The factorization in (32) shows us that the first factor ($t_{i,j} t_{j,i}$) will always be greater than or equal to zero while the second factor can be either positive or negative. Hence, by isolating the second factor and recognizing that trust values equal to 1 produce the smallest possible reduction in the second factor, we can state the condition that guarantees the potential for two agents to form a trust-based pair coalition.

Proposition 1: Any two agents $i, j \in N$ will never form a trust-based pair coalition if $\sigma_{i,j} < \lambda_{i,j} + \lambda_{j,i}$. Otherwise, the potential exists for agent i and j to form a trust-based pair coalition.

Proposition 2: If two agents can never form a trust-based pair coalition, then the best strategy for both agents is to never trust each other (i.e. $t_{i,j} = t_{j,i} = 0$).

In general, proposition 1 does not extend to trust-based coalitions larger than two due to the complex coupling of trust dynamics between different agents as coalitions grow larger. For example, two agents who may produce a negative trust payoff value as a pair may actually realize a positive trust payoff with the addition of a third agent. This situation occurs if both agents have positive trust relationships with the third agent that outweighs their own negative trust relationship. Such a situation is common in real world scenarios, and justifies the importance of various trusted third parties, such as escrow companies, website authentication services, and couples therapists.

In light of this, we can mathematically justify a condition similar to proposition 1 that is valid for coalitions of any size – but only for a special type of trust game.

Theorem 3: A trust-based coalition $S \subseteq N$ will never form if:

$$\sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j} < \sum_{\substack{i,j \in S \\ \forall i \neq j}} \frac{\lambda_{i,j}}{t_{j,i}} \quad (33)$$

$$\{\forall i, j \in S: t_{i,j} t_{j,i} = k\}$$

Proof: Let $S \subseteq N$ and $t_{i,j}t_{j,i} = k$ for all $i, j \in S$. Then, by substituting k into (31):

$$v(S) = \sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j} k - \sum_{\substack{i,j \in S \\ \forall i \neq j}} \frac{\lambda_{i,j}}{t_{j,i}} k \quad (34)$$

$$v(S) = k \left(\sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j} - \sum_{\substack{i,j \in S \\ \forall i \neq j}} \frac{\lambda_{i,j}}{t_{j,i}} \right) \quad (35)$$

Because k is a constant that is always greater than or equal to zero, we can clearly see that the second factor affects whether or not $v(S)$ is positive or negative. Hence, if the second term in the second factor is larger than the first term, then a coalition S will never form. This completes the proof.

4.4 Incorporating Context into a Trust Game

In practice, trust is often defined relative to some context. Context allows individuals to simplify complex decision-making scenarios by focusing on more narrow perspectives of situations or others, avoiding the potential for inconvenient paradoxes.

Coalitional trust games can also be defined relative to different contexts using the multi-issue representation²⁶, where we use the words “context” and “issue” interchangeably.

Definition: A *multi-issue representation* is composed of a collection of coalitional games, each known as an issue, $(N_1, v_1), (N_2, v_2), \dots, (N_k, v_k)$, which together constitute the coalitional game (N, v) where

- $N = N_1 \cup N_2 \cup \dots \cup N_k$
- For each coalition $S \subseteq N$, $v(S) = \sum_{i=1}^k v_i(S \cap N_i)$

This approach allows us to define an arbitrarily complex trust game that can be easily decomposed into simpler trust games relative to a particular context. A set of agents in one context can overlap partially or complete with another set of agents in another context. And one can choose to treat the coalitional game in one big context, or the union of any number of contexts based on some decision criteria.

5. APPLYING A TRUST GAME TO UNMANNED COOPERATIVE CONTROL

In this section, we apply a coalitional trust game in a specific unmanned cooperative control application: the unmanned military convoy. Currently, the United States Army Tank Automotive Research, Development, and Engineering Center is funding the Convoy Active Safety Technology (CAST) program, which aims to improve convoy operations with the installation of a small kit in the cab of a tactical vehicle²⁷. The kit connects actuators to the steering wheel, gas pedal, and brake pedal, and uses various sensors, such as RADAR, LIDAR, and electro-optical/infrared cameras, to sense a vehicle’s environment and safely drive the vehicle. Our goal with the trust game is to understand how trust-based coalitions will form under different attitudes of trust in a convoy scenario.

We begin with a simple convoy scenario that models a four-vehicle convoy, $N = \{1,2,3,4\}$, which intends to move together in a single file. The value of each index into N represents the vehicle’s position in the convoy. For this scenario, we interpret the trust synergy in coalition to represent the vehicles in the coalition moving forward. Thus, we set the values in the trust synergy matrix Σ equal to the number of vehicles that will move forward if the two vehicles are moving forward (inclusive of the two vehicles). We interpret the trust liability in coalition to represent the vulnerability of vehicles in the coalition to stop moving. Thus, we set the values in the trust liability matrix



Tactical wheeled vehicles equipped with Convoy Active Safety Technology (CAST)

Λ equal to the number of vehicles can prevent a particular vehicle from moving forward in a vehicle coalition pair. Note that in a more realistic unmanned convoy trust game, the values in Σ and Λ could be based on additional factors, such as the presence of hostile forces, the smoothness of the road, the time of day, weather conditions, vehicle reliability, or mission importance.

Definition: The values in Σ and Λ for a 4-convoy trust game are:

$$\Sigma = \begin{bmatrix} 0 & 2 & 3 & 4 \\ 2 & 0 & 3 & 4 \\ 3 & 3 & 0 & 4 \\ 4 & 4 & 4 & 0 \end{bmatrix} \quad \Lambda = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 2 & 2 & 0 & 2 \\ 3 & 3 & 3 & 0 \end{bmatrix} \quad (36)$$

First, let us analyze this game as an additive trust game. While there are infinitely many solutions for T that conform to (4), the most obvious solution is the extreme situation where no vehicle trusts any other vehicle – or, when T is the identity matrix ($T = I$). In this case, it can clearly be seen from (31) that no vehicle will ever affect another vehicle, either positively or negatively. Thus, each vehicle will ultimately form a singleton coalition and fail to work cooperatively with any other vehicle.

Next, let us analyze another extreme situation where every vehicle completely trusts every other vehicle – or, when $T = [1]_{4 \times 4}$. As such, we can enumerate the trust payoff values for each possible coalition.

$$\begin{aligned} v(\{1,2\}) &= 1; & v(\{1,3\}) &= 1; & v(\{1,4\}) &= 1; & v(\{2,3\}) &= 0; & v(\{2,4\}) &= 0; & v(\{3,4\}) &= -1; \\ v(\{1,2,3\}) &= 2; & v(\{1,2,4\}) &= 2; & v(\{1,3,4\}) &= 1; & v(\{2,3,4\}) &= -1; & v(\{1,2,3,4\}) &= 2; \end{aligned} \quad (37)$$

The results in (37) provide us an interesting insight, in that all vehicles behind the lead vehicle find higher values of trust payoff with the lead vehicle than with the nearest vehicle. As such, as long as the lead vehicle is a member of a trust-based coalition in this game, there will be no incentive for any other vehicle to abandon the coalition. Thus, the vehicles ultimately form the grand coalition. Note, however, that the formation of a grand coalition does not imply that the trust game is superadditive or convex. This assertion is justified with the observation that $v(\{3,4\}) \neq v(\{3\}) + v(\{4\}) = 0$.

In order to form a convex 4-convoy trust game, we must satisfy the conditions in (25), which ensure that all trust payoff values in any coalition are at least large as any sub-coalition. While there are infinitely many solutions for T that conform to (25), the games with the highest trust payoff have either one of the following trust matrices:

$$T_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad T_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \quad T_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad T_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \quad (38)$$

T_1, T_2, T_3 , and T_4 are modified versions of $[1]_{4 \times 4}$ and all produce the same results in the trust payoff value function. The main modification ensures that vehicles 3 and 4 have no trust toward each other since the trust liabilities between them always outweigh their trust synergies. The following is the enumeration of the trust payoff values for the 4-convoy trust game with the highest trust payoff:

$$\begin{aligned} v(\{1,2\}) &= 1; & v(\{1,3\}) &= 1; & v(\{1,4\}) &= 1; & v(\{2,3\}) &= 0; & v(\{2,4\}) &= 0; & v(\{3,4\}) &= 0; \\ v(\{1,2,3\}) &= 2; & v(\{1,2,4\}) &= 2; & v(\{1,3,4\}) &= 2; & v(\{2,3,4\}) &= 0; & v(\{1,2,3,4\}) &= 3; \end{aligned} \quad (39)$$

The deep insight we gain from analyzing (38) and the results in (39) is that all vehicles behind the lead vehicle need only trust the lead vehicle in the convoy to move forward, provided the lead vehicle trusts every other vehicle to follow it. This echoes the intuition seen in Jean-Jacques Rousseau's classic "stag hunt" game, where there is no incentive for any player to cheat by not cooperating as long as each player can trust others to do the same. This type of game differs from the well-known "prisoners' dilemma" game, where the dominant strategy is to confess and not cooperate with any other player²⁸.

Personal experience suggests that drivers in convoy-like traffic patterns rarely place large amounts of trust in neighboring drivers, as our model corroborates. In the event a driver becomes stuck in a giant traffic jam, he likely will not feel betrayed by the driver directly in front. Instead, he will unconsciously begin gauging the coalitional value of the traffic jam by considering his level of trust in the lead driver in the traffic jam, whether in visible range or not. In most cases, the driver monitors the traffic flow or listens to traffic reports to gauge his trust for the lead driver. He may also unconsciously consider other drivers in the traffic jam and estimate their trust perceptions of the traffic jam to gauge the coalition's value. In the event a driver cannot accurately gauge the value of the traffic jam, he may choose to leave the traffic jam and attempt to join another traffic coalition with a higher payoff value. These types of driver behaviors are generally not performed when the trust for the lead driver to move forward is high. Yet, these behaviors feel necessary when the trust lessens since they attempt to resolve coalitional or environmental uncertainties.

For unmanned military convoys, our results suggest that follower vehicles need only communicate with the lead vehicle to ensure trustworthy cooperation. The lead vehicle needs only to broadcast pertinent information to the followers, and the followers need only to acknowledge receiving the information from the lead to signal agreement. This hub-and-spoke communications network would therefore foster the reciprocity necessary to cultivate trust between the leader and its followers while also keeping the computational complexity of the network to a minimum of $O(n)$. The presence of the solution T_4 suggests that trust-based redundancy can be achieved with the second vehicle in the event of a catastrophic failure to the lead vehicle. The cost of the trust-based redundancy would require an additional $(|N| - 2)$ point-to-point connections, but the computational complexity would not change.

We conclude this section by generalizing the convoy trust game for any number of vehicles and prove the solution for the highest payoff trust-based coalition. Our proof shows that all vehicles behind the lead vehicle in a convoy need only trust the lead vehicle, and no other vehicle, to move forward so long as the lead vehicle trusts every other vehicle to follow it.

Definition: The values in Σ and Λ for a convoy trust game with $|N|$ vehicles are:

$$\Sigma = [\sigma_{i,j}]_{|N| \times |N|} = \begin{cases} \sigma_{i,j} = 0, & i = j \\ \sigma_{i,j} = \max(\{i, j\}), & i \neq j \end{cases} \quad (40)$$

$$\Lambda = [\lambda_{i,j}]_{|N| \times |N|} = \begin{cases} \lambda_{i,j} = 0, & i = j \\ \lambda_{i,j} = i - 1, & i \neq j \end{cases} \quad (41)$$

Theorem 4: The convoy trust game that produces the grand coalition with highest payoff value has a trust matrix that conforms to the following construction:

$$T = [t_{i,j}]_{|N| \times |N|} = \begin{cases} t_{i,j} = 1, & i = j \\ t_{i,j} = 1, & i \neq j, \min(\{i, j\}) = 1 \\ t_{i,j} = t_{j,i} \in \{0, 1\}, & i \neq j, \min(\{i, j\}) = 2 \\ t_{i,j} = 0, & i \neq j, \min(\{i, j\}) > 2 \end{cases} \quad (42)$$

Proof:

Suppose we generalize the values in Σ and Λ according to (40) and (41), respectively. According to proposition 1, two agents $i, j \in N$ will never form a trust-based coalition pair if $\sigma_{i,j} < \lambda_{i,j} + \lambda_{j,i}$. Thus, by substitution:

$$\max(\{i, j\}) < (i - 1) + (j - 1) \quad (43)$$

$$\max(\{i, j\}) < i + j - 2 \quad (44)$$

We see that if i is the maximum value, then $0 < j - 2$. Similarly, if j is the maximum value, then $0 < i - 2$. Thus, the inequality in (44) tells us that any vehicle behind the second vehicle will never form a trust-based coalition with any other vehicle behind the second vehicle. Therefore, by proposition 2, the best strategy for these vehicles is to have no trust for each other; hence $t_{i,j} = 0$ when $\min(\{i, j\}) > 2$ for $i \neq j$.

Since the result in (44) implies that trust-based coalition formation is possible with the lead vehicle and the second vehicle, we must analyze the trust payoff values for coalitions with these vehicles. Using (32) and our definitions in (40) and (41), the trust payoff values for a coalition in the convoy trust game is:

$$v(S) = \sum_{\substack{i,j \in S \\ \forall i > j}} t_{i,j} t_{j,i} \left(\max(\{i, j\}) - \frac{i-1}{t_{j,i}} - \frac{j-1}{t_{i,j}} \right) \quad (45)$$

From (45), we define trust payoff values for any pair of vehicles as:

$$v(\{i, j\}) = t_{i,j} t_{j,i} \left(\max(\{i, j\}) - \frac{i-1}{t_{j,i}} - \frac{j-1}{t_{i,j}} \right) \quad (46)$$

Let us first analyze coalition formation with the lead vehicle. If $i = 1$, then $\max(\{i, j\}) = j$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{1, j\}) = t_{1,j} t_{j,1} \left(j - \frac{j-1}{t_{1,j}} \right) \quad (47)$$

$$v(\{1, j\}) = j t_{1,j} t_{j,1} - j t_{j,1} + t_{j,1} \quad (48)$$

$$v(\{1, j\}) = t_{j,1} (j t_{1,j} - j + 1) \quad (49)$$

The result in (49) show that the highest trust payoff value is achieved when both the lead vehicle and any other vehicle completely trust each other (i.e., when $t_{1,j} = t_{j,1} = 1$). However, to justify this assertion, we must also show this is true when $j = 1$. If $j = 1$, then $\max(\{i, j\}) = i$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{i, 1\}) = t_{i,1} t_{1,i} \left(i - \frac{i-1}{t_{1,i}} \right) \quad (50)$$

$$v(\{i, 1\}) = i t_{i,1} t_{1,i} - i t_{i,1} + t_{i,1} \quad (51)$$

$$v(\{i, 1\}) = t_{i,1} (i t_{1,i} - i + 1) \quad (52)$$

Both (49) and (52) confirm that the highest trust payoff is achieved when both the lead vehicle and any other vehicle completely trust each other. Therefore, $t_{i,j} = 1$ when the $\min(\{i, j\}) = 1$ for $i \neq j$.

Now, we analyze coalition formation with the second vehicle. If $i = 2$, then $\max(\{i, j\}) = j$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{2, j\}) = t_{2,j} t_{j,2} \left(j - \frac{1}{t_{j,2}} - \frac{j-1}{t_{2,j}} \right) \quad (53)$$

$$v(\{2, j\}) = t_{2,j} t_{j,2} j - t_{2,j} - j t_{j,2} + t_{j,2} \quad (54)$$

$$v(\{2, j\}) = t_{j,2} (j t_{2,j} - j + 1) - t_{2,j} \quad (55)$$

The highest trust payoff that can be achieved with the second vehicle is equal to zero, and this only occurs when both vehicles either have complete trust in each other (i.e., when $t_{2,j} = t_{j,2} = 1$) or no trust in each other (i.e., when $t_{2,j} = t_{j,2} = 0$). Any other combination of trust values will produce negative trust payoff values. However, to justify this

assertion, we must also show this is true when $j = 2$. If $j = 2$, then $\max(\{i, j\}) = i$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{i, 2\}) = t_{i,2}t_{2,i} \left(i - \frac{i-1}{t_{2,i}} - \frac{1}{t_{i,2}} \right) \quad (56)$$

$$v(\{i, 2\}) = it_{i,2}t_{2,i} - it_{i,2} + t_{i,2} - t_{2,i} \quad (57)$$

$$v(\{i, 2\}) = t_{i,2}(it_{2,i} - i + 1) - t_{2,i} \quad (58)$$

Both (55) and (58) confirm that the highest trust payoff that can be achieved with the second vehicle is equal to zero. Therefore, $t_{i,j} = t_{j,i} \in \{0,1\}$ when $\min(\{i, j\}) = 2$ for $i \neq j$. To complete the proof, we simply state our assumption that each vehicle fully trusts itself, since it is impossible for a vehicle to diverge from a singleton coalition. Therefore, $t_{i,j} = 1$ when $i = j$. This completes the proof.

REFERENCES

- [1] Luhmann, N., "Familiarity, Confidence, Trust: Problems and Alternatives," In D. Gambetta (Ed.) [Trust: Making and Breaking Cooperative Relations], Basil Blackwell, New York, 94-108 (1988).
- [2] Singer, P.W., "Advanced Warfare: How We Might Fight With Robots," [Wired For War], Penguin Press, New York, NY, (2009).
- [3] Adams, B. & Webb, R., "Trust in Small Military Teams," 7th International Command and Control Technology Symposium, (2002).
- [4] Ramchurn, S.D., Huynh, D. & Jennings, N.R., "Trust in Multi-Agent Systems", The Knowledge Engineering Review 19(1), 1-25 (2004).
- [5] Sen S., "Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents," In Proc 2nd International Conference on Multi-Agent Systems, AAAI Press, Menlo Park, CA, 322-329, (1996).
- [6] Mukherjee R., Banerjee B., and Sen S., "Learning Mutual Trust," In R. Falcone, M. Singh, and Yao-Hua Tan (Eds.) [Trust in Cyber-societies], Springer-Verlag, Berlin Heidelberg, 145-158, (2001).
- [7] Castelfranchi, C. and Falcone, R., "Principles of trust for MAS: cognitive anatomy, social importance, and quantification," In Proc International Conference of Multi-Agent Systems (ICMAS98), 72-79, (1998).
- [8] Castelfranchi C. and Falcone R., "Social trust: A cognitive approach," In C. Castelfranchi and Yao-Hua Tan (Eds.) [Trust and Deception in Virtual Societies], Kluwer Academic Publishers, 55-90, (2000).
- [9] Castelfranchi, C. and Falcone, R., "Trust is much more than subjective probability: Mental components and sources of trust," In Proc 33rd Hawaii International Conference on System Sciences (online edition), 6, IEEE Computer Society, (2000).
- [10] Witkowski M., Artikis A., Pitt J., "Experiments in building experiential trust in a society of objective-trust based agents" In R. Falcone, M. Singh, and Yao-Hua Tan (Eds.) [Trust in Cyber-Societies], Springer-Verlag, Berlin Heidelberg, 111-132, (2001).
- [11] Sabater J. and Sierra C., "REGRET: a reputation model for gregarious societies," In C. Castelfranchi and L. Johnson (Eds.) [Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems], 475-482, (2002).
- [12] Teacy, L., Patel J., Jennings, N., Luck M., "Coping with Inaccurate Reputation Sources: Experimental Analysis of a Probabilistic Trust Model," In Proc. 4th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS), ACM Press, 997-1004, (2005).
- [13] Theodorakopoulos, G. and Baras, J., "In Trust Models and Trust Evaluation Metrics for Ad Hoc Networks," IEEE Journal on Selected Areas in Communications, 24(2), 318-328, (2006).
- [14] Wang Y. and Singh M., "Formal Trust Model for Multiagent Systems," In Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI), 1551-1556, (2007).

- [15] Abdul-Rahman, A. and Hailes, S., "Supporting Trust in Virtual Communities," In Proc. 33rd Hawaii International Conference on Systems Science, (2000).
- [16] Zhang, W., Das, S., Liu, Y. "A Trust Based Framework for Secure Data Aggregation in Wireless Sensor Networks," In Proc. 3rd IEEE Communication Society of Sensor and Ad Hoc Communications and Networks, Reston, VA, 60-69, (2006).
- [17] Ballal, P. and Lewis, F., "Trust-Based Collaborative Control for Teams in Communication Networks," In Proc. 26th Army Science Conference. Orlando, FL, (2008).
- [18] Parkes, D. and Shneidman, J., "Distributed Implementations of Vickrey-Clarke-Groves Mechanisms," In Proc. 3rd International Joint Conference on Autonomous Agents and MultiAgent Systems, 261-268, (2004).
- [19] Zacharia G. and Maes P., "Trust through reputation mechanisms," Applied Artificial Intelligence, 14, 881–907, (2000).
- [20] Poslad S., Calisti M. , Charlton P., "Specifying standard security mechanisms in multi-agent systems," In Workshop on Deception, Fraud and Trust in Agent Societies, AAMAS 2002, Bologna, Italy, 122–127, (2002).
- [21] "FIPA Abstract Architecture Specification," Foundation For Intelligent Physical Agents, <http://fipa.org/specs/fipa00001/>, (2002)
- [22] Shoham, Y. and Leyton-Brown, K., "Teams of Selfish Agents: An Introduction to Coalitional Game Theory," In Y. Shoham and K. Leyton-Brown [Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations], Cambridge University Press, Cambridge, 367-391, (2009).
- [23] Baras, J., Jiang, T., Purkayastha, P., "Constrained Coalitional Games and Networks of Autonomous Agents," ISCCSP 2008, Malta, 972-979, (2008).
- [24] Branzei, R., Dimitrov, D., Tijs, S., "A new characterization of convex games", Discussion Paper 2004-109, Tiburg University, Center of Economic Research, (2004).
- [25] Lovasz, L., "Submodular function and convexity," In A. Bachem, M. Grotschel, and B. Korte (ed.) [Mathematical Programming. The State of the Art], 235-257, Springer-Verlag, (1983).
- [26] Conitzer, V. and Sandholm, T. "Computing Shapley values, manipulating value division schemes, and checking core membership in multi-issue domains," In Proc. AAAI Conference on Artificial Intelligence, (2004).
- [27] Schoenherr, E., "Moving Future Convoy Operations with Convoy Active Safety Technologies (CAST)," TARDEC Accelerate Magazine, 4, 74-78, (2009).
- [28] Dixit, A. and Nalebuff B., "Prisoners' Dilemmas and How to Resolve Them," [The Art of Strategy], W.W. Norton and Company, New York, 64-101, (2008).